《抽樣方法》

<table>
<tr><td>試題評析</td><td>今年考題具鑑別度，只要把握好計算時間，拿到分數並不困難。程度好的考生可拿80~90分左右，一般考生則落在60~70之間。
第一題：爲群集隨機抽樣考題，均等機率與非均等機率之比較，在歷屆考題中皆曾出現過。
第二題：爲傳統基本觀念，在歷屆考題中亦曾出現過。
第三題：爲分層隨機抽樣考題，比例配置與紐門配置之計算，只需把握好計算時間。
第四題：爲群集隨機抽樣考題，於歷屆考題中亦曾出現過。
第五題：爲傳統基本考題，簡單隨機抽樣與系統隨機抽樣之比較，只需把握兩方法差異之關鍵即可拿分。</td></tr>
<tr><td>高分命中</td><td>第一題：《高點抽樣方法講義第二回》，馮國經編撰，頁19-20，考題7與考題8。
《高點抽樣方法講義第二回》，馮國經編撰，頁47，考題1。
第二題：《高點抽樣方法講義第一回》，馮國經編撰，頁41，中央極限定理。
《高點抽樣方法講義第一回》，馮國經編撰，頁42，(1)母體平均值之估計。
《高點抽樣方法講義第一回》，馮國經編撰，頁45，(丙)抽取不放回，母體變異數已知 $Var\left(\bar{y}\right)=\left(1-f\right)\frac{S^2}{n}$ 之證明。
第三題：《高點抽樣方法講義第一回》，馮國經編撰，頁111-112，考題16。
《高點抽樣方法講義第一回》，馮國經編撰，頁82-83，考題1。
《高點抽樣方法講義第一回》，馮國經編撰，頁15，考題23。
第四題：《高點抽樣方法講義第二回》，馮國經編撰，頁51，考題6。
第五題：《高點抽樣方法講義第二回》，馮國經編撰，頁81-82，[系統隨機抽樣法與簡單隨機抽樣法之比較說明]。</td></tr>
</table>

本試題可能使用之標準常態值如下：

$Z_{0.025}=1.96$，$Z_{0.05}=1.645$，$Z_{0.1}=1.28$

一、研究者想估計，A城之成年市民年所得（單位：百萬元），該市依街道等因素分成53個里，且已知A城之成年市民總數為3,500人。因經費及時間因素只能隨機抽出5個里作調查，調查結果如下表：

樣本	1	2	3	4	5
成年市民總數	80	125	42	50	67
成年市民年收入總數	9.6	12.1	4.2	6.5	5.2

(一)試用二種方式來估計A城之成年市民年所得。（10分）

(二)請依估計的準確度為選擇之依據，何種方式較佳？並請說明。（12分）

答：

(一)

1.群集隨機抽樣法

$$\bar{y}_t=\frac{Y_1+Y_2+Y_3+Y_4+Y_5+}{5}=\frac{9.6+12.1+4.2+6.5+5.2}{5}=7.52$$

$$\hat{Y}_{cl}=N\cdot\bar{y}_t=53\cdot 7.52=398.56$$

2.比例機率抽樣法

$$\hat{Y}_{ppz}=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{Z_i}=\frac{M}{n}\sum_{i=1}^{n}\frac{Y_i}{M_i}$$
$$=\frac{3500}{5}\left(\frac{9.6}{80}+\frac{12.1}{125}+\frac{4.2}{42}+\frac{6.5}{50}+\frac{5.2}{67}\right)$$
$$=700\cdot 0.5244=367.08$$

(二)

1.群集隨機抽樣法

$$\operatorname{var}(\hat{Y}_{cl})=N^2\cdot\operatorname{var}(\bar{y}_t)=N^2\cdot(1-f)\frac{s_e^2}{n}$$
$$=53^2\cdot(1-\frac{5}{53})\cdot\frac{10.687}{5}=5437.5456$$
$$s_e^2=\frac{\sum_{i=1}^{n}(Y_i-\bar{y}_t)^2}{n-1}=\frac{1}{n-1}\left[\sum Y_i^2-\frac{(\sum Y_i)^2}{n}\right]$$
$$=\frac{1}{5-1}\left[325.5-\frac{1413.76}{5}\right]=10.687$$

2.比例機率抽樣法

$$\operatorname{var}(\hat{Y}_{ppz})=\frac{M^2\sum_{i=1}^{n}(\bar{Y}_i-\bar{\bar{y}}_t)^2}{n(n-1)}=\frac{3500^2\cdot 0.001692}{5\cdot(5-1)}=1036.35$$
$$\sum_{i=1}^{5}(\bar{Y}_i-\bar{\bar{y}}_t)^2=\sum\bar{Y}_i^2-\frac{(\sum\bar{Y}_i)^2}{n}=0.056692-\frac{(0.5244)^2}{5}$$
$$=0.056692-0.055$$
$$=0.001692$$

3.因 $\operatorname{var}(\hat{Y}_{ppz})<\operatorname{var}(\hat{Y}_{cl})$ ，所以比例機率法較佳。

二、若 $Y_1,Y_2...,Y_n$ 係自總數為N，平均數為μ，變異數為 σ^2 之母體，以不放回之隨機抽樣方法抽出的隨機樣本，且 $\bar{Y}=\sum_{i=1}^{n}Y_i/n \quad \forall n\geq 30$ ，試問：

(一) $\bar{Y}$ 是否為μ之不偏估計量？其分配為何？（8分）

(二)Var($\bar{Y}$)之不偏估計量為何？並說明所使用符號之意義。（5分）

答：

(一)

$$1.\overline{Y}=\frac{\sum_{i=1}^{n}Y_i}{n}$$

$$=\frac{1}{n}[E(Y_1)+E(Y_2)+\cdots+E(Y_n)]$$

$$=\frac{1}{n}[u+u+\cdots+u]$$

$$=\frac{1}{n}\cdot n\cdot u=u$$

2.依中央極限定理

$$\overline{Y}=\frac{\sum_{i=}^{n}Y_i}{n}\to\sim N(u,\frac{6^2}{n})$$

(二)依抽取不放回，可得

$$Var(\overline{Y})=(1-f)\frac{S^2}{n}=\left(\frac{N-n}{N}\right)\cdot\frac{S^2}{n}=\left(\frac{N-n}{N-1}\right)\cdot\frac{\sigma^2}{n}$$

1. f 爲抽出率。
2. $\frac{N-n}{N}$ 與 $\frac{N-n}{N-1}$ 爲有限母體校正因子。
3. S^2 爲母體變異數。
4. n 爲抽出樣本數。

三、精英國中想要了解國二學生的英文閱讀能力，將國二學生依其英文考試成績高低分成三組，各組人數依序分別為55、80及65人，並隨機抽出50位同學為樣本，請問：

(一)採比例抽樣法各組應抽取幾位？(5分)

(二)測驗後各組成績結果如下表，試求國二學生英文閱讀能力的平均分數及變異數？(12分)

組別	1	2	3
測驗總分	1116	1295	599
測驗分數的變異數	105.14	158.20	186.13

(三)學校擬在學期結束前，再做一次測驗。若各組選取每位學生作測驗的成本都相同，且應試學生為60位，則各組應抽取多少位學生作測驗？(12分)

答：

(一)比例配置

$$n_h=\frac{N_h}{\sum_{h=1}^{L}N_h}\cdot n$$

$$n_1=\frac{55}{200}\cdot 50=13.75\cong 14$$

$$n_2=\frac{80}{200}\cdot 50=20$$

$$n_3=\frac{65}{200}\cdot 50=16.25\cong 16$$

(二)

組別	1	2	3
N_h	55	80	65
n_h	14	20	16
總分	1116	1295	599
$\bar{y}_h$	79.71	64.75	37.44
s_h^2	105.14	158.2	186.13

1. $$\bar{y}_{st}=\sum_{h=1}^{3}W_h\bar{y}_h=\frac{55}{200}\cdot 79.71+\frac{80}{200}\cdot 64.75+\frac{65}{200}\cdot 37.44$$
$$=59.99$$

2. $$\operatorname{var}(\bar{y}_{st})=\sum_{h=1}^{3}W_h^2\cdot\operatorname{var}(\bar{y}_h)=\sum_{h=1}^{3}\frac{N_h^2}{N^2}\cdot(1-f_h)\frac{s_h^2}{n_h}=2.2989$$

	$\frac{N_h^2}{N^2}\cdot(1-f_h)\cdot\frac{s_h^2}{n_h}$
N_1	$\frac{55^2}{200^2}\cdot(1-\frac{14}{55})\frac{105.14}{14}=0.4234$
N_2	$\frac{80^2}{200^2}\cdot(1-\frac{20}{80})\frac{158.2}{20}=0.9492$
N_3	$\frac{65^2}{200^2}\cdot(1-\frac{16}{65})\frac{186.13}{16}=0.9263$
	$\sum_{h=1}^{3}\frac{N_h^2}{N^2}\cdot(1-f_h)\frac{s_h^2}{n_h}=2.2989$

(三)各組成本相同，$E\left(s_i^2\right)=S_i^2$，採紐門配置。

$$n_h=\frac{N_hS_h}{\sum N_hS_h}\cdot n$$

$$n_1=\frac{55\times 10.25}{2456.75}\cdot 60=13.76\cong 14$$

$$n_2=\frac{80\times 12.58}{2456.75}\cdot 60=24.57\cong 24$$

$$n_2=\frac{65\times 13.64}{2456.75}\cdot 60=21.65\cong 22$$

四、A工廠生產電子產品的電路板，每50塊裝成一箱，每塊有35個晶片。品管人員由生產完成之12箱中隨機抽出一箱，再由該箱中隨機抽出10塊，檢測後發現晶片有缺陷的個數分別是3、2、0、1、4、3、0、0、1、2。試估計該箱中有缺陷的晶片總個數，並求變異數？若品管人員再隨機抽一箱電路板做檢測，希望缺陷晶片總個數的容忍誤差為30個且信賴度為90%，請問從該箱電路板中，應抽出幾塊做檢驗？（24分）

答：

(一)

1. $\overline{A_t}=\dfrac{A_1+\cdots A_{10}}{10}=\dfrac{3+2+\cdots+2}{10}=\dfrac{16}{10}=1.6$

$\hat{A}_{cl}=N\cdot\overline{A_t}=50\cdot 1.6=80$

2. $\operatorname{var}(\hat{A}_{cl})=N^2\cdot(1-f)\cdot\dfrac{s_e^2}{n}$

$$=50^2\cdot(1-\frac{10}{50})\cdot\frac{20.44}{10}=408.8$$

$$s_e^2=\frac{\sum_{i=1}^{n}(A_i-\overline{A_t})^2}{n-1}$$

$$=\frac{1}{n-1}\left[\sum A_i^2-\frac{(\sum A_i)^2}{n}\right]$$

$$=\frac{1}{9}\left[44-\frac{16^2}{10}\right]$$

$$=2.044$$

(二)

$$n_0=\left(\frac{Z\cdot N\cdot s_e}{B}\right)^2=\left(\frac{1.645^2\cdot 50^2\cdot 2.044}{30^2}\right)=15.36$$

$$n=\frac{15.36}{1+\frac{15.36}{50}}=\frac{15.36}{1.3072}=11.75\cong 12$$

五、若分別採用系統抽樣法與簡單隨機抽樣法所得樣本資訊估計母體平均數μ作估計，請依母體資料與其次序之情況，對估計量之特性（如不偏性，相對有效性……），比較說明何者較佳？為什麼？（12分）

答：

(一)環境說明，參考資料來源：黃文隆教授，《抽樣方法》，滄海書局、儲全滋教授，《抽樣方法》，三民書局。

SST(總變異)=SSC(組間變異)+SSE(組內變異)

$$SST=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij}-\overline{Y}\right)^2=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij}-\overline{y}_{i.}\right)^2+\sum_{i=1}^{k}\sum_{i=1}^{n}\left(\overline{y}_{i.}-\overline{Y}\right)$$

$$1.\ SST=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij}-\overline{Y}\right)^2$$

$$2.\ SSC=\sum_{i=1}^{k}\sum_{i=1}^{n}\left(\overline{y}_{i.}-\overline{Y}\right)$$

$$3.\ SSE=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij}-\overline{y}_{i.}\right)^2$$

(二)定理說明一數學推導過程

$$\sum_{i=1}^{K}\sum_{j=1}^{n}\left(y_{ij}-\overline{Y}\right)^2=\sum_{i=1}^{K}\sum_{j=1}^{n}\left(y_{ij}-\overline{y}_{i\cdot}\right)^2+\sum_{i=1}^{K}\left(\overline{y}_{i\cdot}-\overline{Y}\right)^2\times n$$

$$(N-1)\cdot S^2=\sum_{i=1}^{K}\sum_{j=1}^{n}\left(y_{ij}-\overline{y}_{i\cdot}\right)^2+\sum_{i=1}^{K}\left(\overline{y}_{i\cdot}-\overline{Y}\right)^2\times n$$

$$(N-1)\cdot S^2=SSE+\sum_{i=1}^{K}\left(\overline{y}_{i\cdot}-\overline{Y}\right)^2\times\frac{1}{K}\times K\times n$$

$$(N-1)\cdot S^2=SSE+n\cdot k\cdot Var\left(\overline{y}_{sy}\right)$$

$$n\cdot k\cdot V\mathrm{ar}\left(\overline{y}_{sy}\right)=(N-1)\cdot S^2-SSE$$

$$\mathrm{Var}\left(\overline{y}_{sy}\right)=\frac{N-1}{n\cdot k}\cdot S^2-\frac{SSE}{n\cdot k}$$

$$\mathrm{Var}\left(\overline{y}_{sy}\right)=\frac{N-1}{N}\cdot S^2-\frac{SSE}{N}\ldots\ldots\ldots\ldots\ldots(1)$$

$$\mathrm{Var}\left(\overline{y}\right)=(1-f)\frac{S^2}{n}\ldots\ldots\ldots\ldots\ldots(2)$$

$$(2)-(1)$$

$$\mathrm{Var}\left(\overline{y}\right)-Var\left(\overline{y}_{sy}\right)=(1-f)\frac{S^2}{n}-\frac{N-1}{N}S^2+\frac{SSE}{N}=\frac{SSE}{N}+\left(1-\frac{n}{N}\right)\frac{S^2}{n}-\frac{N-1}{N}S^2$$

$$=\frac{SSE}{N}+\frac{S^2}{n}-\frac{S^2}{N}-\left(1-\frac{1}{N}\right)\cdot S^2=\frac{SSE}{N}+\left(\frac{1}{n}-\frac{1}{N}-1+\frac{1}{N}\right)\cdot S^2$$

$$=\frac{SSE}{N}+\left(\frac{1}{n}-1\right)\cdot S^2=\frac{SSE}{N}-\left(1-\frac{1}{n}\right)\cdot S^2=\frac{SSE}{N}-\left(\frac{n-1}{n}\right)\cdot S^2$$

$$=\frac{K(n-1)}{n\cdot K}\cdot\frac{SSW}{K(n-1)}-\frac{K\cdot(n-1)}{n\cdot K}\cdot S^2=\frac{K(n-1)}{n\cdot K}\left(\frac{SSE}{K(n-1)}-S^2\right)$$

$$=\frac{K(n-1)}{n\cdot K}\left(MSE-S^2\right)$$

(三)定理說明—結論：$Var(\bar{y})-Var(\bar{y}_{sy})=\dfrac{k(n-1)}{nk}\left[MSE-S^2\right]$

1.若 $MSE<S^2$，則系統隨機抽樣法之樣本平均變異數相對大於簡單隨機抽樣法之樣本平均變異數；即是系統估計量的準確度較簡單估計量為低。

2.若 $MSE=S^2$，則系統隨機抽樣法之樣本平均變異數等於簡單隨機抽樣法之樣本平均變異數；即是系統估計量的準確度與簡單估計量相同。

3.若 $MSE>S^2$，則系統隨機抽樣法之樣本平均變異數相對小於簡單隨機抽樣法之樣本平均變異數；即是系統估計量的準確度較簡單估計量為高。

(四)定理說明—深入說明：

1.在系統隨機抽樣的過程中，系統樣本平均數的變異數可表示成

$$Var(\bar{y}_{sy})=\frac{N-1}{N}S^2-\frac{SSE}{N}$$ 。

2.當系統樣本內的總變異(組內變異)越大時 → SSE就越大 → 系統樣本平均值的變異數就越小。

3.主要是因為系統隨機抽樣所抽出的樣本可以均勻分布，所以樣本內的變數資料差異就越大。

(五)簡單隨機抽樣與系統隨機抽樣之估計量不偏性說明：

簡單隨機抽樣與系統隨機抽樣之母體平均數 $\bar{Y}$ 的估計量皆具備不偏性。